

CLASIFICADORES TURBO

Elizabeth Tapia, José C. González

Luis J. García-Villalba

Departamento de Ingeniería Telemática
Universidad Politécnica de Madrid
{etapia, jcg}@gsi.dit.upm.es

Departamento de Ciencias de la Computación
Universidad Complutense de Madrid
javiergv@sip.ucm.es

ABSTRACT

By means of turbo coding models, learning in M -output domains is formulated in terms of binary learning methods. Likewise well known adaptive binary boosting methods, learning formulation makes explicit use of binary learner's performance. In addition, the underlying coding structure allows the introduction of memory, which contributes to enhance the overall learning performance.

1. INTRODUCCIÓN

En esta comunicación se estudia el diseño de algoritmos de aprendizaje supervisado de complejidad reducida orientados a la solución de problemas de clasificación múltiple. El tratamiento de la complejidad inherente a valores elevados del número M de clases se afronta a través de una generalización de la estrategia ECOC basada en el empleo de códigos Turbo. La formulación ECOC permite la descomposición del problema de clasificación múltiple de orden M en n problemas de clasificación binaria, donde n la longitud de palabra código Turbo. La hipótesis final es el resultado de un proceso de decodificación turbo sobre la una versión inmersa en ruido (de aprendizaje) de una palabra código turbo.

2. CÓDIGOS TURBO

Los códigos turbo [1] son una instancia particular de códigos recursivos [2] basados en el empleo de códigos componentes convolucionales [3] recursivos, sistemáticos y de complejidad reducida. En su forma más sencilla (código turbo de tasa $R = \frac{1}{3}$), la transmisión de una secuencia de k bits de información, requiere la transmisión de una palabra código turbo de longitud $n = 3 \cdot (k + B)$, donde B es la memoria del código convolucional que define el código turbo. Una diferencia fundamental entre los modelos de codificación recursivos y los modelos clásicos de codificación [4] (códigos de bloque, convolucionales, BCH, etc.) es la existencia de algoritmos recursivos de decodificación. El modelado de códigos recursivos a través de los denominados grafos de Tanner, permite la interpretación del proceso de decodificación global a través algoritmos de pasajes de mensajes [6][7] similares al algoritmo de Pearl [8] para redes bayesianas.

3. MODELO ECOC

Sea $C : X \rightarrow Y$, $|Y| = M$ una clase objetivo y sea $c \in C$ un concepto objetivo para el aprendizaje supervisado mediante una muestra de ejemplos S , $|S| = m$. Sea Γ un código de cardinalidad M y longitud n de palabra código. En su forma estándar, el algoritmo ECOC [9] establece una transformación Θ uno a uno, entre clases y palabras códigos. Cuando Θ se aplica sobre las instancias en la muestra S , su resultado es la definición de un vector de muestras binarias $\mathbf{S} = [S_i]$, $1 \leq i \leq n$. Las muestras binarias S_i resultantes pueden utilizarse entonces para entrenar clasificadores binarios WL_i , $1 \leq i \leq n$. Tal conjunto de clasificadores binarios posee en general una capacidad limitada de aprendizaje cuyo valor se refleja en la probabilidad de error correspondiente e_i , $1 \leq i \leq n$. Sea D_i una distribución de probabilidad definida sobre elementos de la muestra S_i , entonces

$$e_i = P_{S_i \sim D_i} [WL_i(\mathbf{x}) \neq y] \quad (1)$$

La explotación de las características de error de cada uno de los clasificadores binarios WL_i , $1 \leq i \leq n$ es una característica deseable en el diseño de algoritmos de clasificación, por cuanto su efecto más inmediato suele ser una mejora considerable en la capacidad global de aprendizaje. Este efecto es claramente observable en los algoritmos de boosting adaptativos [10] como AdaBoost. La aplicación de modelos de codificación clásicos en la estrategia ECOC descarta el uso de modelos adaptativos de clasificación. Este hecho queda evidenciado en el modelo de predicción ECOC: dado un vector de atributos $\mathbf{x} \in X$ el proceso de asignación de un valor de clase requiere la evaluación del conjunto de n clasificadores binarios para calcular

$$\mathbf{r} = [WL_1(\mathbf{x}), \dots, WL_n(\mathbf{x})] \quad (2)$$

seguido de un proceso de decodificación por distancia mínima de Hamming entre la palabra código inmersa en ruido (de aprendizaje) y el conjunto de palabras código Γ . Finalmente, la correspondencia entre clases y palabras códigos establecida por Θ permite la realización del proceso de inferencia. Un problema fundamental en la estrategia ECOC es la complejidad creciente tanto para valores elevados de M como para la mejora de su capacidad de generalización, dado que en ambos casos se requiere un incremento de la longitud de palabra código.

4. MODELO RECOG

De forma alternativa, consideremos una extensión del modelo ECOC basado en códigos recursivos. Sea Γ_F una instancia particular de los mismos aplicada en la descomposición ECOC. La clasificación de un vector de atributos $\mathbf{x} \in X$ requiere una vez más la evaluación del conjunto de los n clasificadores binarios a los efectos de generar una versión inmersa en ruido \mathbf{r} de la palabra código transmitida. Sin embargo, en esta ocasión y debido a la estructura de codificación recursiva subyacente, la palabra código recibida define implícitamente un conjunto $\mathbf{r}_i, 1 \leq i \leq F-1$, de palabras código inmersas en ruido, donde F es el número de códigos componentes que definen Γ_F . La decodificación iterativa a partir de la palabra código recibida \mathbf{r} requiere el conocimiento de la estadística del canal sobre cada uno de los símbolos de palabra código. Asumiendo la transmisión de palabras código sobre un canal aditivo, discreto y sin memoria [11] o de forma equivalente predicciones binarias independientes con "ruido de aprendizaje" binario y aditivo., las estimaciones de los errores e_i en cada una de las muestras $S_i, 1 \leq i \leq n$, constituyen una primera aproximación a la estadística de un modelo de canal y son suficientes para la aplicación de una estrategia de decodificación recursiva. De aquí en adelante nos referimos a esta nueva estrategia de aprendizaje como el modelo RECOG (Recursive ECOC). Sus características de adaptación y escalabilidad están fundadas en un modelo potente de codificación como son los códigos recursivos.

En esta comunicación se exponen resultados experimentales del modelo RECOG para la instancia más conocida de códigos recursivos, los códigos turbo. El modelo resultante se identificará bajo la denominación RECOG_Turbo.

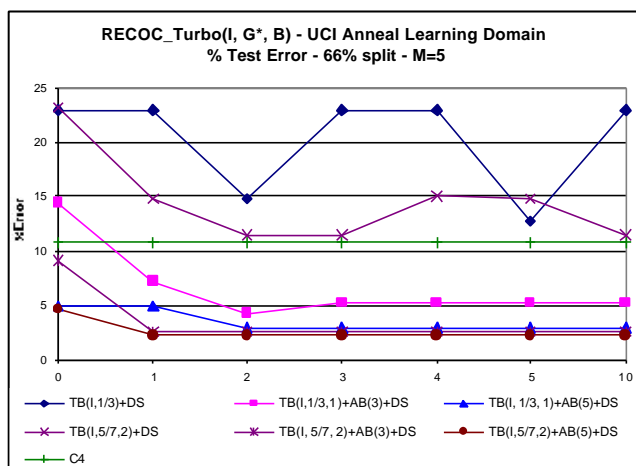


Figura 1. Ensayos con RECOG_Turbo de tasa $R = \frac{1}{3}$

En la figura 1, se muestran gráficas representativas de los comportamientos observados en diversos dominios de aprendizaje. En el caso indicado, se trata con $M=5$ clases y se analiza el valor del error en el conjunto de test (se ha usado 66% de los datos en entrenamiento y resto de test). Las curvas corresponden al algoritmo RECOG_Turbo (siglas TB) para un número de

iteraciones I sobre un código turbo construido con códigos convolucionales de una matriz generadora $G=[1 \ G^*]$ (notación octal) con memoria B . Puede observarse la mejora con el incremento de B (1-2), así como la sensible mejora cuando se combina con boosting interno $AB(T)$ sobre la base de T iteraciones. En todos los casos el algoritmo base usado es el Decisión Stumps (DS). A los fines de comparación se indica el comportamiento del algoritmo C4 (árboles de decisión).

5. CONCLUSIONES

En esta comunicación se resumen las primeras conclusiones sobre la investigación y desarrollo de algoritmos de clasificación de complejidad reducida para problemas de clasificación múltiple con características de adaptabilidad y escalabilidad (con potenciales aplicaciones en problemas de biología y de procesamiento de información). Los resultados obtenidos señalan una importante vía de investigación.

6. REFERENCIAS

- [1] Berrou C. and Glavieux A., "Near Optimum Error Correcting Coding and Decoding: Turbo Codes", IEEE Trans. on Communications, Vol. 44, No. 10, pp. 1261-1271, 1996.
- [2] Tanner, M., A recursive approach to Low Complexity Error Correcting Codes. IEEE Trans. Inf. Theory, Vol. 27, pp. 533- 547, 1981.
- [3] Viterbi A. J., "Convolutional codes and Their Performance in Communication Systems", IEEE Trans. on Communications Technology, Vol. COM 19, No 5, pp. 751- 771, 1971.
- [4] Blahut R., *Error Control Codes*, Addison Wesley 1984.
- [5] McEliece R.J., MacKay D., "Turbo Decoding as an Instance of Pearl's Belief Propagation Algorithm", IEEE Journal on Sel. Areas in Communications, Vol. 16, No 2, pp. 140-152, 1998.
- [6] Kschischang F., Frey, B., Iterative decoding of compound codes by probability propagation in graphical models, IEEE Journal on Sel. Areas in Communications Vol. 16, No. 2, pp. 219-230, 1998.
- [7] Hagenauer J., "Iterative Decoding of Binary Block and Convolutional Codes", IEEE Trans. on Inf. Theory, Vol. 42, No 2, pp. 429-445, 1996.
- [8] Pearl, J., Fusion, Propagation and Structuring in Belief Networks, Artificial Intelligence, Vol. 29, pp. 241-268, 1986.
- [9] Dietterich, T., Bakiri G., Error-correcting output codes: A general method for improving multiclass inductive learning programs, Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91), pp. 572-577, Anaheim, CA: AAAI Press, 1991.
- [10] Schapire R., Y. Singer, "Improved Boosting Algorithms Using Confidence - Rated Predictions", Machine Learning, Vol. 37, No. 3, pp. 277 -296, 1999.
- [11] Csiszár I., Körner J, *Information Theory: Coding Theorems for Discrete Memoryless Systems*, Academic Press, Inc. (London) LTD, 1981.